THE UNIVERSITY of EDINBURGH

Abstract for the "Dealing with Data Conference" (26th August 2014)

Title: Exploring challenges and strategies when using medical databases in research - a review
Authors: Natalia Calanzani[1], Gaby Vojt[1], Christine Campbell[1], David Weller[1]

[1] University of Edinburgh, Centre for Population Health Sciences, Medical School

Abstract:

Background: The use of medical databases in research is well-established in the UK, and is likely to intensify further with the launch of the *care.data* programme in England. However, managing and analysing these datasets is not straightforward. Specialised knowledge is required, in addition to computer resources and the development of comprehensive data analysis plans. A lot can be learnt from the published literature regarding the shortcomings of these data and possible strategies to deal with them. This presentation will discuss the findings from a literature review (ongoing) which aims to investigate such challenges and strategies in order to inform a large study using linked patient databases. This review was also presented at the 43rd Annual Scientific Meeting of the Society for Academic Primary Care on the 11th July 2014 (original abstract available on page 2).

Methods: The presentation will outline several identified challenges, which were organised according to different required steps to obtain and analyse medical data. Whenever available, strategies to deal with these will also be presented. The presentation will be a snapshot of the review findings, focusing on aspects which are more relevant to the Conference.

Results: In order to reach a wider audience, the presentation will not emphasise findings on challenges or solutions regarding specific statistical methods. Instead, it will focus on findings such as the large sizes and complexities of the datasets, or difficulties when trying to identify medical conditions when several codes are used for the same diagnosis. Likewise, corresponding strategies will be related to suggestions regarding data management, such as the need to create coding lists or metadata files.

Implications: It is expected that, in combination with the introduction of the Research Data Management support service, the audience will identify several ways in which this support could be helpful for anyone analysing medical databases. This could include, for example, the provision of storage solutions, help with the development of a data management plan or the importance of creating metadata files.

Natalia Calanzani
Research Fellow
Email: natalia.calanzani@ed.ac.uk

THE UNIVERSITY of EDINBURGH

ORIGINAL ABSTRACT
(PRESENTED AT THE SAPC MEETING ON THE 11TH JULY 2014)
http://www.sapc.ac.uk/index.php/conf2014

Abstract No. 0228
Title Exploring limitations and solutions when using medical databases in primary care research: a literature review

Abstract
The problem: Databases containing patient routine data are widely used in primary care research. From late 2014, anonymised GP records will be stored in a centralised database as part of the England NHS Care.data programme; these data will be available for researchers. In order to accurately interpret analyses, limitations of the databases should be acknowledged, and ways to overcome these limitations better understood. This review describes several such limitations, focusing on databases provided by the Clinical Practice Research Datalink (CPRD, a well-established UK provider of primary care data) and on understanding Read codes (the standard clinical terminology system used in general practice, adopted by the CPRD). This review also presents evidence based strategies on how to address these challenges.

The approach: The purpose of this review was to inform an ongoing exploratory study that uses CPRD data to analyse patterns of consultation from non-responders to colorectal cancer screening. We conducted literature searches (in medical and technology databases) in October and November 2013, checked the CPRD publications list at the CPRD website and the reference lists of included studies. We included 1) studies discussing challenges when using medical databases (including those not specific to CPRD but with results applicable to CPRD); and/or 2) studies discussing the limitations of using Read Codes; and/or 3) studies outlining approaches to overcome limitations of using these data. Narrative synthesis was used to describe the findings.

Findings: Studies described inter-practice and inter-person variations in coding, as daily clinical practice is complex and there can be multiple choices to code a single diagnosis. The existence of financial incentives can also add to coding variability. Recording certain conditions or medical/socio-demographic characteristics such as diabetes, comorbidities and ethnicity can be problematic. The validity of diagnoses can vary according to different conditions. Epidemiological challenges included biases, possibility of residual confounding and the likelihood of obtaining statistically significant results with large sample sizes. Approaches to overcome limitations varied from general systematic procedures, to strategies to address specific limitations. Procedures included prior assessment of whether the database is appropriate to answer the research questions and several discrete steps when preparing data for analysis. Additional validation analyses can be carried out by researchers. The need to understand clinical contexts was emphasised by several authors.

Consequences: There are several limitations when using medical databases, but there is also substantial evidence-based information on how to address these. Limitations should be taken into account from the initial stages of research design. Results from this review can help researchers using both CPRD data/analysing Read codes, and those utilising similar medical databases. In light of the launching of the C are.data programme, a better understanding of the use of medical data for research is even more important.

Affiliations (1) University of Edinburgh, Edinburgh, UK

Authors Natalia Calanzani (1) Presenting
Gaby Vojt (1)
Christine Campbell (1)
David Weller (1)

Natalia Calanzani
Research Fellow
Email: natalia.calanzani@ed.ac.uk